Optimal design for multivariate multiple linear regression with set-identified response

Qiyu Li[a,*], Ilya Molchanov[b,*]

[a] *University of Bern, Institute of Mathematical Statistics and Actuarial Science, Sidlerstrasse 5, CH-3012 Bern, Switzerland and Swiss Group for Clinical Cancer Research (SAKK), Effingerstrasse 35, CH-3008 Bern, Switzerland*

[b] *University of Bern, Institute of Mathematical Statistics and Actuarial Science, Sidlerstrasse 5, CH-3012 Bern, Switzerland*

Abstract

We consider the partially identified regression model with set-identified responses, where the estimator is the set of the least square estimators obtained for all possible choices of points sampled from set-valued observations. We address the issue of determining the optimal design for this case and show that, for objective functions mimicking those for several classical optimal designs, their set-identified analogues coincide with the optimal designs for point-identified real-valued responses.

Keywords: design, partially identified model, random convex set, regression, set-identified response

1. Introduction

Consider the basic regression model

$$y_i = x_i^\top \theta + \varepsilon_i\,, \quad i = 1, \dots, n\,,$$

where the design points $x_1, \dots, x_n$ belong to $\mathbb{R}^{r+1}$ called the *design space*, y_i, $i = 1, \dots, n$, are observed real-valued responses, θ is a vector of $(r+1)$ unknown numerical parameters, and $\varepsilon_1, \dots, \varepsilon_n$ are independent identically distributed (i.i.d.) centred random variables with variance $\mathrm{Var}(\varepsilon_i) = \sigma^2$. This setting includes the classical multivariate linear model, and also other models, like the quadratic one that appears if

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i1}^2.$$

The basic problem in the theory of optimal design for regression models aims to identify the locations of design points $x_1, \dots, x_n$ which ensure the best properties of the unbiased

*Corresponding author

Email addresses: qiyu.li@stat.unibe.ch (Qiyu Li), ilya.molchanov@stat.unibe.ch (Ilya Molchanov)

estimator $\hat{\theta}$ of θ. As the objective function to minimize, one can choose, e.g., the sum of the variances of the components of $\hat{\theta}$ (the criterion function for the A-optimal design) or the largest variance of $a^\top \hat{\theta}$ over all unit vectors a (which yields the E-optimal design). Further optimality criteria lead to a multitude of other optimal designs, see [1, 14].

In this paper we consider the situation when the possibly multivariate response y is set-identified, so instead of observing $y_1, \dots, y_n$, the statistician is only given sets $Y_1, \dots, Y_n$ that contain the true observations. It is assumed that the specific points $y_i \in Y_i$ are chosen by a completely unknown selection mechanism which is not a subject to statistical modelling. In this *partially identified* setting, it is not possible to come up with a single-valued estimator for θ. We follow the approach advocated by Beresteanu and Molinari [2] who suggested considering all possible points (selections) $y_i \in Y_i$, $i = 1, \dots, n$, fitting to them the linear regression model in order to obtain particular (least squares) estimator $\hat{\theta}$ and, finally, use the set of all estimators $\hat{\theta}$ obtained in this way as the estimator for the set-identified regression, see also [13]. The most important special case arises if the observations $Y_1, \dots, Y_n$ are intervals on the line; then one talks about interval regression, see also [3, 5] for an alternative approach based on the interval arithmetics. The main reason of having interval-identified data are variability and uncertainty. For example, the temperature on a certain day is typically reported by weather forecasts as an interval between the lowest and the highest temperatures. This interval represents the variability of the temperature. In social surveys, salaries of respondents are usually reported as intervals. Another example in the field of oncology is the time to recurrence of a tumor. The recurrence status of a patient is assessed by imaging techniques such as a CT scan at every visit, which is not scheduled every day but rather every two or three months. Therefore, we only know that recurrence occurs between two visits but not its exact time point. In this case, the data of time to recurrence are also interval-identified. In case of several interval-identified responses, the obtained multiple response is set-identified by a parallelepiped or its subset determined by the imposed constraints.

In this paper, we address the issue of optimal design in the partially identified least squares setting of [2]. The crucial issue is to properly handle the variance of the estimated parameters; unlike the expectation, the variance of random sets is rather poorly understood, see [12].

In Section 2 we introduce the notation used throughout the paper and recall some definitions and results from random set theory. This is followed by Section 3, where we recall the classical A-, G- and E-optimal designs with point-identified data. In Section 4, we introduce the objective functions for the set-identified setting and prove that the corresponding optimal designs coincide with the classical A-, G- and E-optimal design under some assumptions on the model structure. As a corollary, we deduce that the A-, G-, and E-optimal multiresponse designs in the multiresponse point-identified setting coincide with their classical analogues; this extends the result of Chang [4] derived for D-optimal designs.

2. Random convex sets and their expectation

We use $\|\cdot\|$ to denote the Euclidean norm. Let $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d : \|v\| = 1\}$ denote the unit sphere in $\mathbb{R}^d$. If $d = 1$, then the sphere consists of two points $\{-1, 1\}$. The family of non-empty compact convex sets (also called convex bodies) in $\mathbb{R}^d$ is denoted by $\mathcal{K}(\mathbb{R}^d)$. The *support function* of $K \in \mathcal{K}(\mathbb{R}^d)$ is defined as

$$s(K, v) = \max_{y \in K} v^\top y, \qquad v \in \mathbb{S}^{d-1},$$

so that $s(K, v)$ is the signed length of the projection of K onto the line with direction v. If $K = [a, b]$, then $s(K, 1) = b$ and $s(K, -1) = -a$.

The support function identifies uniquely the corresponding convex compact set and satisfies

$$\begin{aligned} s(tK, v) &= ts(K, v), \quad t > 0, \\ s(-K, v) &= s(K, -v), \\ s(K_1 + K_2, v) &= s(K_1, v) + s(K_2, v), \end{aligned}$$

where $-K = \{-x : \ x \in K\}$ is the centrally symmetric set to K, and

$$K_1 + K_2 = \{x + y : \ x \in K_1, y \in K_2\}$$

is the Minkowski sum of two convex bodies K_1 and K_2.

Let $(\Omega, \mathfrak{F}, \mathbf{P})$ be a nonatomic probability space, where all random vectors and random sets are defined. The map $\boldsymbol{Y} : \Omega \mapsto \mathcal{K}(\mathbb{R}^d)$ is called a random convex body, if $\{\omega \in \Omega : \boldsymbol{Y}(\omega) \cap A \neq \emptyset\} \in \mathfrak{F}$ for every compact set A in $\mathbb{R}^d$. A random vector $\boldsymbol{y}$ in $\mathbb{R}^d$ is called a *selection* of $\boldsymbol{Y}$ if $\boldsymbol{y}(\omega) \in \boldsymbol{Y}(\omega)$ for almost all $\omega \in \Omega$. We denote this as $\boldsymbol{y} \in \boldsymbol{Y}$ a.s.

We assume throughout that $\boldsymbol{Y}$ is *integrably bounded*, that is, $\|\boldsymbol{Y}\| = \sup\{\|\boldsymbol{y}\| : \ \boldsymbol{y} \in \boldsymbol{Y}\}$ is an integrable random variable. In this case, all selections of $\boldsymbol{Y}$ are integrable and the *expectation* $\mathbf{E}\boldsymbol{Y}$ is defined as the set of $\mathbf{E}\boldsymbol{y}$ for all selections $\boldsymbol{y}$ of $\boldsymbol{Y}$. Equivalently, $\mathbf{E}\boldsymbol{Y}$ is the convex body that satisfies

$$\mathbf{E}s(\boldsymbol{Y}, v) = s(\mathbf{E}\boldsymbol{Y}, v), \quad v \in \mathbb{S}^{d-1}.$$

If $\boldsymbol{Y} = [\boldsymbol{y}_{\mathrm{L}}, \boldsymbol{y}_{\mathrm{U}}]$ is the interval then $\mathbf{E}\boldsymbol{Y} = [\mathbf{E}\boldsymbol{y}_{\mathrm{L}}, \mathbf{E}\boldsymbol{y}_{\mathrm{U}}]$.

Similarly, the conditional expectation of $\boldsymbol{Y}$ given a random vector (or matrix) $\boldsymbol{x}$ is defined as

$$\mathbf{E}(\boldsymbol{Y}|\boldsymbol{x}) = \big\{\mathbf{E}(\boldsymbol{y}|\boldsymbol{x}) : \boldsymbol{y} \in \boldsymbol{Y} \text{ a.s.}\big\},$$

and then $\mathbf{E}(s(\boldsymbol{Y}, v)|\boldsymbol{x}) = s(\mathbf{E}(\boldsymbol{Y}|\boldsymbol{x}), v)$ a.s.

3. Classical optimal designs in the multiresponse setting

Consider i.i.d. sample (x_i, y_i), $i = 1, \dots, n$, where $y_i = (y_{i1}, \dots, y_{ip})^\top \in \mathbb{R}^p$ designates response and $x_i = (1, x_{i1}, \dots x_{ir})^\top \in \mathbb{R}^{r+1}$ is the vector composed of explanatory variables.

Let

$$\mathscr{X} = (x_1^\top, \dots, x_n^\top)^\top = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1r} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nr} \end{pmatrix}$$

be the *design matrix* with n rows and $r+1$ columns. Collect all the responses in the matrix

$$\mathscr{Y} = (y_1^\top, \dots, y_n^\top)^\top = \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix}.$$

Consider the regression model

$$\mathscr{Y} = \mathscr{X}\Theta + \mathscr{E},$$

where

$$\Theta = \begin{pmatrix} \theta_{01} & \cdots & \theta_{0p} \\ \theta_{11} & \cdots & \theta_{1p} \\ \vdots & \vdots & \vdots \\ \theta_{r1} & \cdots & \theta_{rp} \end{pmatrix}$$

is the matrix of unknown parameters, and the matrix $\mathscr{E} = (\varepsilon_1^\top, \dots, \varepsilon_n^\top)^\top$ consists of i.i.d. square integrable centred random vectors $\varepsilon_i = (\varepsilon_{i1}, \dots \varepsilon_{ip})^\top$ such that $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) = \sigma_{jk}$ for $i = 1, \dots, n$ and $j, k = 1, \dots, p$ and $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'k}) = 0$ for $i \neq i'$.

Assume that the model has full rank, meaning that

$$\Sigma = \mathscr{X}^\top \mathscr{X}$$

is invertible. We stress that Σ depends on $\mathscr{X}$. The least square estimator of Θ is

$$\hat{\Theta} = \Sigma^{-1}\mathscr{X}^\top\mathscr{Y}.$$

Then $\mathbf{E}\hat{\Theta} = \Theta$ and

$$\mathrm{Cov}(\hat{\Theta}_{(j)}, \hat{\Theta}_{(k)}) = \sigma_{jk}\Sigma^{-1}, \quad j, k = 1, \dots, p,$$

where $\hat{\Theta}_{(k)}$ denotes the kth column of $\hat{\Theta}$.

The sum of variances of all elements in the matrix $\hat{\Theta}$ is

$$\sum_{k=1}^{p} \sigma_{kk}\,\mathrm{Tr}(\Sigma^{-1}),$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. The A-optimal design minimizes this sum of variances and so can be obtained by minimizing $\mathrm{Tr}(\Sigma^{-1})$ over all designs. Depending on the framework, the designs mentioned here could be exact n-trial designs or approximate ones. Note that the objective function for the A-optimality does not depend on the dimension of the response variable. The objective function can be equivalently written as

$$\sum_{j=1}^{r+1} \frac{1}{\lambda_j},$$

where $\lambda_1, \ldots, \lambda_{r+1}$ are the eigenvalues of Σ.

The E-optimal design is chosen to minimize the variance of the least well estimated contrast $a^\top(\hat{\Theta}_{(1)}, \ldots, \hat{\Theta}_{(p)})^\top$ under the constraint $\|a\| = a^\top a = 1$. This objective function could be expressed as the maximum element on the diagonal of Σ^{-1}, which is also known as the MV-optimal designed introduced by Jacroux [7]. The E-optimality criterion can be equivalently expressed as minimization of $\max(\lambda_1^{-1}, \ldots, \lambda_{r+1}^{-1})$.

The variance of the response at certain $x \in \{1\} \times \mathbb{R}^r$ could be expressed as

$$\operatorname{Var}(\hat{\boldsymbol{y}}(x)) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} x^\top \Sigma^{-1} x.$$

The design which minimizes the maximum of the variance of the predicted response over an arbitrary design region $\mathcal{I} \subset \{1\} \times \mathbb{R}^r$ is called G-optimal. The corresponding objective function is $\max_{x \in \mathcal{I}} x^\top \Sigma^{-1} x$.

4. Optimal designs for set-identified response

Assume that the response is set-identified, and the statistician observes compact convex sets $Y_1, \ldots, Y_n$ in $\mathbb{R}^p$, where possible responses $y_1, \ldots, y_n$ take their values. The explanatory variables are assumed to be point-identified. Following [2] and given the i.i.d. data $(x_i, Y_i)_{i=1}^n$, the least square estimators of the regression coefficients Θ form the *family of matrices*

$$\hat{\boldsymbol{\Theta}} = (\mathscr{X}^\top \mathscr{X})^{-1} \mathscr{X}^\top \left\{ \begin{pmatrix} y_1^\top \\ \vdots \\ y_n^\top \end{pmatrix} : y_i \in Y_i \right\} = \Sigma^{-1} \sum_{i=1}^n \operatorname{diag}(x_i) \boldsymbol{G}_i, \tag{1}$$

where throughout this paper $\operatorname{diag}(\cdot)$ of a vector denotes the diagonal matrix built from this vector, and $\boldsymbol{G}_i$ is the set of $(r+1) \times p$ matrices with

$$\boldsymbol{G}_i = \left\{ \begin{pmatrix} y_i^\top \\ \vdots \\ y_i^\top \end{pmatrix} : y_i \in Y_i \right\}, \quad i = 1, \ldots, n.$$

Denote by $\mathbf{E}_{\mathscr{X}}$ the expectation assuming that the design matrix is $\mathscr{X}$. Note that $\hat{\boldsymbol{\Theta}}$ is a set of matrices, each of them is a least square estimator for a certain sample of responses $y_1, \ldots, y_n$ arbitrarily selected from $Y_1, \ldots, Y_n$. In order to define its variance, we consider products of all matrices with a given $u \in \mathbb{S}^{p-1}$; and then the support function of the obtained random convex set in $\mathbb{R}^{r+1}$ in direction v from the unit sphere $\mathbb{S}^r$ in $\mathbb{R}^{r+1}$. In other words, we work with the variance

$$\operatorname{Var}_{\mathscr{X}} s(\hat{\mathbf{\Theta}}u, v) = \mathbf{E}_{\mathscr{X}}(s(\hat{\mathbf{\Theta}}u, v) - s(\mathbf{E}_{\mathscr{X}}(\hat{\mathbf{\Theta}}u), v))^2$$

of the support function of $\hat{\mathbf{\Theta}}u$ and aim to minimize it as function of the design. Note that $\hat{\mathbf{\Theta}}u$ is a random convex set in $\mathbb{R}^{r+1}$, and its expectation is defined in Section 2.

Following the classical definitions of A-, G- and E-optimal designs, we define the objective function for these designs in the set-identified framework as

$$f^A(\mathscr{X}) = \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^r} \operatorname{Var}_{\mathscr{X}} s(\hat{\mathbf{\Theta}}u, v)\, dv du, \tag{2}$$

$$f^G(\mathscr{X}) = \max_{x\in\mathcal{I}} \int_{\mathbb{S}^{p-1}} \operatorname{Var}_{\mathscr{X}} s(\hat{\mathbf{\Theta}}^\top x, u)\, du = \max_{x\in\mathcal{I}} \int_{\mathbb{S}^{p-1}} \operatorname{Var}_{\mathscr{X}} s(\hat{\mathbf{\Theta}}u, x)\, du, \tag{3}$$

$$f^E(\mathscr{X}) = \max_{u\in\mathbb{S}^{p-1}} \max_{v\in\mathbb{S}^r} \operatorname{Var}_{\mathscr{X}} s(\hat{\mathbf{\Theta}}u, v). \tag{4}$$

Here the integrals over spheres are understood with respect to a finite rotation invariant measure (the Haar measure) and $\mathcal{I}$ is a compact subset of $\{1\} \times \mathbb{R}^r$.

Example 4.1 (Univariate set-identified response). If $p = 1$ and $\boldsymbol{Y} = [\boldsymbol{y}_{\mathrm{L}}, \boldsymbol{y}_{\mathrm{U}}]$ (like in the examples mentioned in the introduction), then $\hat{\mathbf{\Theta}}$ is a family of $(r+1)\times 1$ matrices, equivalently, vectors in $\mathbb{R}^{r+1}$. In this case, $u = \pm 1$ and the integrals (or maximum) with respect to u in the objective functions reduce to the sum (or maximum) over $u = \pm 1$ of the variances of the support function of the predicted response at $v \in \mathbb{S}^r$.

We denote $M(x) = \mathbf{E}(\boldsymbol{Y}|\boldsymbol{x} = x)$ and $m(x) = \mathbf{E}(\boldsymbol{y}|\boldsymbol{x} = x)$, and also $M_i = M(x_i)$ and $m_i = m(x_i)$.

Theorem 4.2. *Assume that*

$$s(\boldsymbol{Y}, u) - s(M(\boldsymbol{x}), u) = \varepsilon(u), \quad u \in \mathbb{S}^{p-1}, \tag{5}$$

where ε is a random function on the unit sphere that does not depend on $\boldsymbol{x}$ and satisfies $\mathbf{E}\varepsilon(u) = 0$ and $\operatorname{Var}(\varepsilon(u)) = \sigma_u^2 < \infty$ for all $u \in \mathbb{S}^{p-1}$. Then the designs minimizing the objective functions defined in (2) *and* (3) *correspond to the classical A- and G-optimal designs.*

Remark 4.3. Condition (5) is a modelling assumption. In case of random intervals $\boldsymbol{Y} = [\boldsymbol{y}_{\mathrm{L}}, \boldsymbol{y}_{\mathrm{U}}]$, it means that

$$\begin{cases} \boldsymbol{y}_{\mathrm{L}} = \mathbf{E}(\boldsymbol{y}_{\mathrm{L}}|\boldsymbol{x}) - \varepsilon(-1), \\ \boldsymbol{y}_{\mathrm{U}} = \mathbf{E}(\boldsymbol{y}_{\mathrm{U}}|\boldsymbol{x}) + \varepsilon(1), \end{cases}$$

for a centred random vector $(\varepsilon(-1), \varepsilon(1))$ such that

$$\varepsilon(1) + \varepsilon(-1) \geq \mathbf{E}(\boldsymbol{y}_{\mathrm{L}}|\boldsymbol{x}) - \mathbf{E}(\boldsymbol{y}_{\mathrm{U}}|\boldsymbol{x}) \quad \text{a.s.}$$

The latter condition replicates the requirement that $\mathbf{P}(\boldsymbol{y}_{\mathrm{U}} > \boldsymbol{y}_{\mathrm{L}}) = 1$.

Proof of Theorem 4.2. First, consider the A-optimal design. Using (1) and the additivity of support function as function of convex bodies,

$$
\begin{aligned}
&s(\hat{\mathbf{\Theta}}u, v) - s(\mathbf{E}_{\mathscr{X}}(\hat{\mathbf{\Theta}}u), v) \\
&\qquad = \sum_{i=1}^{n} \left\{ s\left(\Sigma^{-1}\operatorname{diag}(x_i)\boldsymbol{G}_i u, v\right) - s\left(\Sigma^{-1}\operatorname{diag}(x_i)\mathbf{E}_{\mathscr{X}}(\boldsymbol{G}_i u), v\right) \right\} \\
&\qquad = \sum_{i=1}^{n} \left\{ s\left(\boldsymbol{G}_i u, \operatorname{diag}(x_i)\Sigma^{-1}v\right) - s\left(\mathbf{E}_{\mathscr{X}}(\boldsymbol{G}_i u), \operatorname{diag}(x_i)\Sigma^{-1}v\right) \right\}.
\end{aligned}
$$

Denote $\tilde{v}_i = \operatorname{diag}(x_i)\Sigma^{-1}v$ and

$$
\delta_i(u, v) = s\left(\boldsymbol{G}_i u, \tilde{v}_i\right) - s\left(\mathbf{E}_{\mathscr{X}}(\boldsymbol{G}_i u), \tilde{v}_i\right).
$$

Then

$$
\begin{aligned}
f^A(\mathscr{X}) &= \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^r} \mathbf{E}_{\mathscr{X}} \Big(\sum_{i=1}^{n} \delta_i(u, v) \Big)^2 dv du \\
&= \sum_{i=1}^{n} \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^r} \mathbf{E}_{\mathscr{X}} \delta_i^2(u, v) dv du + \sum_{i \neq i'} \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^r} \mathbf{E}_{\mathscr{X}} (\delta_i(u, v)\delta_{i'}(u, v)) dv du.
\end{aligned}
$$

By expressing $\boldsymbol{G}_i u$ and $\mathbf{E}_{\mathscr{X}}(\boldsymbol{G}_i u)$ as

$$
\begin{aligned}
\boldsymbol{G}_i u &= \left\{ \begin{pmatrix} y_i^\top u \\ \vdots \\ y_i^\top u \end{pmatrix} : y_i \in Y_i \right\}, \\
\mathbf{E}_{\mathscr{X}}(\boldsymbol{G}_i u) &= \left\{ \begin{pmatrix} m_i^\top u \\ \vdots \\ m_i^\top u \end{pmatrix} : m_i \in M_i \right\},
\end{aligned}
$$

we have

$$
\delta_i(u, v) = \max_{y_i \in Y_i} \begin{pmatrix} y_i^\top u \\ \vdots \\ y_i^\top u \end{pmatrix}^\top \tilde{v}_i - \max_{m_i \in M_i} \begin{pmatrix} m_i^\top u \\ \vdots \\ m_i^\top u \end{pmatrix}^\top \tilde{v}_i = \max_{y_i \in Y_i} y_i^\top u \tilde{v}_i^\top \boldsymbol{e} - \max_{m_i \in M_i} m_i^\top u \tilde{v}_i^\top \boldsymbol{e},
$$

where $\boldsymbol{e}$ is the $(r+1)$-dimensional vector with all entries equal to one. Then

$$\begin{aligned}
\mathbf{E}_{\mathscr{X}}\delta_i^2(u,v) &= \mathbf{E}_{\mathscr{X}}\Big[\mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\tilde{v}_i^\top \boldsymbol{e}\Big(\max\{y_i^\top u : y_i\in Y_i\} - \max\{m_i^\top u : m_i \in M_i\}\Big) \\
&\qquad + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\tilde{v}_i^\top \boldsymbol{e}\Big(\inf\{y_i^\top u : y_i\in Y_i\} - \inf\{m_i^\top u : m_i \in M_i\}\Big)\Big]^2 \\
&= (\tilde{v}_i^\top \boldsymbol{e})^2 \mathbf{E}_{\mathscr{X}}\Big[\mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\big(s(Y_i,u) - s(M_i,u)\big) + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\big(-s(Y_i,-u) + s(M_i,-u)\big)\Big]^2 \\
&= (\tilde{v}_i^\top \boldsymbol{e})^2 \mathbf{E}_{\mathscr{X}}\Big[\mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\varepsilon_i(u) - \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\varepsilon_i(-u)\Big]^2 \\
&= (\tilde{v}_i^\top \boldsymbol{e})^2 \mathbf{E}_{\mathscr{X}}\Big[\mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\varepsilon_i(u)^2 + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\varepsilon_i(-u)^2\Big] \\
&= (\tilde{v}_i^\top \boldsymbol{e})^2 \Big[\mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\sigma_u^2 + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\sigma_{-u}^2\Big]. \qquad (6)
\end{aligned}$$

Since

$$\begin{aligned}
\mathbf{E}_{\mathscr{X}}\delta_i^2(u,v) + \mathbf{E}_{\mathscr{X}}\delta_i^2(-u,v) &= (\tilde{v}_i^\top \boldsymbol{e})^2\Big[\mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\sigma_u^2 + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\sigma_{-u}^2 + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}\geq 0\}}\sigma_{-u}^2 + \mathbf{1}_{\{\tilde{v}_i^\top \boldsymbol{e}< 0\}}\sigma_u^2\Big] \\
&= (\tilde{v}_i^\top \boldsymbol{e})^2(\sigma_u^2 + \sigma_{-u}^2),
\end{aligned}$$

we have

$$\begin{aligned}
\int_{\mathbb{S}^{p-1}}\int_{\mathbb{S}^r} \mathbf{E}_{\mathscr{X}}\delta_i^2(u,v)\,dvdu &= \frac{1}{2}\int_{\mathbb{S}^{p-1}}\int_{\mathbb{S}^r}\Big(\mathbf{E}_{\mathscr{X}}\delta_i^2(u,v) + \mathbf{E}_{\mathscr{X}}\delta_i^2(-u,v)\Big)\,dvdu \qquad (7)\\
&= \frac{1}{2}\int_{\mathbb{S}^r}\big((\Sigma^{-1}\operatorname{diag}(x_i)\boldsymbol{e})^\top v\big)^2\,dv \int_{\mathbb{S}^{p-1}}(\sigma_u^2 + \sigma_{-u}^2)du \\
&= \frac{1}{2}\|\Sigma^{-1}x_i\|^2\int_{\mathbb{S}^r}\left(w^\top v\right)^2\,dv\int_{\mathbb{S}^{p-1}}(\sigma_u^2+\sigma_{-u}^2)du,
\end{aligned}$$

where $w = \Sigma^{-1}x_i/\|\Sigma^{-1}x_i\|$ is a unit vector in $\mathbb{R}^{r+1}$. Note that

$$\int_{\mathbb{S}^r}(w^\top v)^2\,dv = \mathbf{E}\left(\frac{Z_j^2}{\sum_{j=1}^{r+1}Z_j^2}\right), \qquad (8)$$

where $i\in\{1,\dots,r+1\}$ and $(Z_1,\dots,Z_{r+1})^\top$ is multivariate standard normal.

Taking the sum on the right-hand side of (8) over $j=1,\dots,r+1$ and noticing that this sum is one, we obtain

$$\int_{\mathbb{S}^r}\left(w^\top v\right)^2\,dv = \frac{1}{r+1},$$

whence

$$\int_{\mathbb{S}^{p-1}}\int_{\mathbb{S}^r}\mathbf{E}_{\mathscr{X}}\delta_i^2(u,v)\,dvdu = \frac{1}{2(r+1)}\|\Sigma^{-1}x_i\|^2\int_{\mathbb{S}^{p-1}}(\sigma_u^2+\sigma_{-u}^2)du.$$

Since ε_i are centred i.i.d., we have

$$\mathbf{E}_{\mathscr{X}}(\delta_i(u,v)\delta_j(u,v)) = 0, \quad i\neq j. \qquad (9)$$

Thus,

$$f^A(\mathscr{X}) = \frac{1}{2(r+1)} \sum_{i=1}^{n} \|\Sigma^{-1} x_i\|^2 \int_{\mathbb{S}^{p-1}} (\sigma_u^2 + \sigma_{-u}^2) du, \tag{10}$$

so that the A-optimal design minimizes $\sum_{i=1}^n \|\Sigma^{-1} x_i\|^2$. This sum can be expressed as

$$\begin{aligned}
\sum_{i=1}^{n} \|\Sigma^{-1} x_i\|^2 &= \sum_{i=1}^{n} x_i^\top \Sigma^{-2} x_i = \sum_{i=1}^{n} \operatorname{Tr}\left(x_i^\top \Sigma^{-2} x_i\right) \\
&= \sum_{i=1}^{n} \operatorname{Tr}\left(\Sigma^{-2} x_i x_i^\top\right) = \operatorname{Tr}\left(\sum_{i=1}^{n} \Sigma^{-2} x_i x_i^\top\right) \\
&= \operatorname{Tr}\left(\Sigma^{-2} \sum_{i=1}^{n} x_i x_i^\top\right) = \operatorname{Tr}\left(\Sigma^{-1}\right) = \sum_{j=1}^{r+1} \frac{1}{\lambda_j},
\end{aligned}$$

where λ_j is the jth eigenvalue of Σ.

The design minimizing this expression is exactly the A-optimal one for the case of real-valued responses.

In order to prove the statement concerning the G-optimal design, note that

$$s(\hat{\boldsymbol{\Theta}}^\top x, u) = \sum_{i=1}^{n} s(\boldsymbol{G}_i{}^\top \Sigma^{-1} \operatorname{diag}(x_i) x, u) = \sum_{i=1}^{n} s(\boldsymbol{G}_i u, \operatorname{diag}(x_i) \Sigma^{-1} x),$$

where $\boldsymbol{G}_i{}^\top$ is the family of all transposed matrices from $\boldsymbol{G}_i$. Denote $\xi_i(x) = \operatorname{diag}(x_i)\Sigma^{-1}x$ and

$$\Delta_i(u, x) = s(\boldsymbol{G}_i u, \xi_i(x)) - s(\mathbf{E}_{\mathscr{X}}(\boldsymbol{G}_i u), \xi_i(x)).$$

Then

$$\begin{aligned}
\operatorname{Var}_{\mathscr{X}} s(\hat{\boldsymbol{\Theta}}^\top x, u)\, du &= \mathbf{E}_{\mathscr{X}} \left(\sum_{i=1}^{n} \Delta_i(u, x)\right)^2 du \\
&= \sum_{i=1}^{n} \mathbf{E}_{\mathscr{X}}(\Delta_i^2(u, x))\, du + \sum_{i \neq j} \mathbf{E}_{\mathscr{X}}(\Delta_i(u, x) \Delta_j(u, x))\, du \\
&= \sum_{i=1}^{n} \mathbf{E}_{\mathscr{X}}(\Delta_i^2(u, x))\, du
\end{aligned}$$

by the same argument as in (9). Following the derivation of (6), we get

$$\Delta_i(u, x) = \max_{y_i \in Y_i} y_i^\top u \xi_i(x)^\top \boldsymbol{e} - \max_{m_i \in M_i} m_i^\top u \xi_i(x)^\top \boldsymbol{e}$$

and

$$\mathbf{E}_{\mathscr{X}} \Delta_i^2(u, x) = (\xi_i(x)^\top \boldsymbol{e})^2 [\mathbf{1}_{\{\xi_i(x)^\top \boldsymbol{e} \geq 0\}} \sigma_u^2 + \mathbf{1}_{\{\xi_i(x)^\top \boldsymbol{e} < 0\}} \sigma_{-u}^2].$$

The idea used to obtain (7) is also applicable here, namely,

$$\begin{aligned}\int_{\mathbb{S}^{p-1}} \mathbf{E}_{\mathscr{X}}(\Delta_i^2(u,x))\,du &= \frac{1}{2}\int_{\mathbb{S}^{p-1}} \Big(\mathbf{E}_{\mathscr{X}}\Delta_i^2(u,x) + \mathbf{E}_{\mathscr{X}}\Delta_i^2(-u,x)\Big)\,du \\ &= \frac{1}{2}(\xi_i(x)^\top \boldsymbol{e})^2 \int_{\mathbb{S}^{p-1}} (\sigma_u^2 + \sigma_{-u}^2)\,du.\end{aligned}$$

Therefore,

$$f^G(\mathscr{X}) = \max_{x\in\mathcal{I}} \sum_{i=1}^n (\xi_i(x)^\top \boldsymbol{e})^2 \frac{1}{2}\int_{\mathbb{S}^{p-1}} (\sigma_u^2 + \sigma_{-u}^2)\,du. \tag{11}$$

Furthermore,

$$\begin{aligned}\sum_{i=1}^n (\xi_i(x)^\top \boldsymbol{e})^2 &= \sum_{i=1}^n (x^\top \Sigma^{-1}\operatorname{diag}(x_i)\boldsymbol{e})^2 = \sum_{i=1}^n (x^\top \Sigma^{-1} x_i)^2 \\ &= \sum_{i=1}^n (x_i^\top \Sigma^{-1} x)^2 = \sum_{i=1}^n \left[(x_i^\top \Sigma^{-1} x)^\top x_i^\top \Sigma^{-1} x\right] \\ &= \sum_{i=1}^n \left[x^\top \Sigma^{-1} x_i x_i^\top \Sigma^{-1} x\right] = x^\top \Sigma^{-1}\left(\sum_{i=1}^n x_i x_i^\top\right)\Sigma^{-1} x \\ &= x^\top \Sigma^{-1} x. \end{aligned} \tag{12}$$

Combine (11) and (12) to see that the design minimizing $f^G(\mathscr{X})$ minimizes $\max_{x\in\mathcal{I}} x^\top\Sigma^{-1}x$, which corresponds to the objective function of the classical G-optimal design. □

Remark 4.4. The Equivalence Theorem by Kiefer and Wolfowitz [9] establishes that the approximate design which is G-optimal is also D-optimal in the case of point-identified univariate response. By letting $\mathcal{I}$ be a singleton in (11), it is immediately seen that the classical D-optimal design minimizes $\operatorname{Var}_{\mathscr{X}} s(\hat{\boldsymbol{\Theta}}^\top x, u)$ for each given x.

Now consider the case of E-optimal designs.

Theorem 4.5. *Assume that* (5) *holds with* ε *being a random function that does not depend on* $\boldsymbol{x}$ *and satisfying* $\mathbf{E}\varepsilon(u) = 0$ *and* $\operatorname{Var}(\varepsilon(u)) = \operatorname{Var}(\varepsilon(-u)) = \sigma_u^2 < \infty$ *for all* $u \in \mathbb{S}^{p-1}$. *Then the design minimizing the objective function* (4) *coincides with the classical* E*-optimal design.*

Proof. Similarly to the case of the A-optimal design, equation (6) can be written as

$$\mathbf{E}_{\mathscr{X}}\delta_i^2(u,v) = (\tilde{v}_i^\top \boldsymbol{e})^2\sigma_u^2 \tag{13}$$

due to the assumption that $\operatorname{Var}(\varepsilon(u)) = \operatorname{Var}(\varepsilon(-u)) = \sigma_u^2$. The expression of variance can be simplified by using (13) and (9), so that

$$f^E(\mathscr{X}) = \max_{u\in\mathbb{S}^{p-1}}\max_{v\in\mathbb{S}^r}\sum_{i=1}^n \mathbf{E}_{\mathscr{X}}\delta_i^2(u,v) = \max_{u\in\mathbb{S}^{p-1}}\sigma(u)^2 \max_{v\in\mathbb{S}^r}\sum_{i=1}^n (\tilde{v}_i^\top \boldsymbol{e})^2. \tag{14}$$

Thus, using the same approach of developing (12), the E-optimal design in the set-identified setting minimizes the maximum over $v \in \mathbb{S}^r$ of

$$\sum_{i=1}^{n} (\tilde{v}_i^\top \boldsymbol{e})^2 = v^\top \Sigma^{-1} v.$$

Finally, observe that this maximum is the maximal eigenvalue of Σ^{-1}, that is,

$$\max_{v \in \mathbb{S}^r} v^\top \Sigma^{-1} v = \max_{j \in (1,\dots,r+1)} \frac{1}{\lambda_j}. \qquad \square$$

Remark 4.6. If $\boldsymbol{Y} = \{\boldsymbol{y}\}$ is a singleton in $\mathbb{R}^p$, we are in the situation of the multiresponse design, and (5) holds with $s(\boldsymbol{Y}, u) = y^\top u$ and $\varepsilon(-u) = -\varepsilon(u)$. Then $\hat{\boldsymbol{\Theta}} = \{\hat{\Theta}\}$ is a singleton, and

$$\mathrm{Var}_{\mathscr{X}}\, s(\hat{\Theta} u, v) = \mathbf{E}_{\mathscr{X}}[((\hat{\Theta} - \mathbf{E}_{\mathscr{X}}\hat{\Theta})u)^\top v]^2.$$

For a matrix A,

$$\int_{\mathbb{S}^r} ((Au)^\top v)^2\, dv = c_r \|Au\|^2$$

with a constant c_r depending only on dimension r and $\max_{v \in \mathbb{S}^r}((Au)^\top, v)^2 = \|Au\|^2$. Therefore, the objective functions of these designs in the multiresponse setting are given by

$$\begin{aligned}
f^A(\mathscr{X}) &= c_r \int_{\mathbb{S}^{p-1}} \mathbf{E}_{\mathscr{X}} \|(\hat{\Theta} - \mathbf{E}_{\mathscr{X}}\hat{\Theta})u\|^2 du, \\
f^G(\mathscr{X}) &= c_{p-1} \max_{x \in \mathcal{I}} \mathbf{E}_{\mathscr{X}} \|(\hat{\Theta} - \mathbf{E}_{\mathscr{X}}\hat{\Theta})^\top x\|^2, \\
f^E(\mathscr{X}) &= \max_{u \in \mathbb{S}^{p-1}} \mathbf{E}_{\mathscr{X}} \|(\hat{\Theta} - \mathbf{E}_{\mathscr{X}}\hat{\Theta})u\|^2.
\end{aligned}$$

By Theorem 4.2, the multiresponse A- and G-optimal designs coincide with their univariate response analogues, the same is the case for E-optimal designs, since the condition of Theorem 4.5 is automatically satisfied.

In the case of univariate responses, $\hat{\Theta}$ becomes a vector $\hat{\theta}$, $u = \pm 1$, and so the objective functions f^A and f^G become $\mathbf{E}\|\hat{\theta} - \mathbf{E}\hat{\theta}\|^2$ and f^G is the maximum of $\mathbf{E}((\hat{\theta} - \mathbf{E}\hat{\theta})^\top x)^2$ (up to dimension-dependent constants).

5. Discussion

The choice of objective functions in our setting is explained by the lack of a standard definition of the variance for random sets, see [12]. In the set-identified multiple response setting, the estimated parameter is a (convex) family $\hat{\boldsymbol{\Theta}}$ of matrices, which is a convex set in dimension $(r+1) \times p$. Then $s(\hat{\boldsymbol{\Theta}} u, v)$ can be interpreted as the support function of $\hat{\boldsymbol{\Theta}}$ in direction $u \otimes v$ understood in the tensor space $\mathbb{R}^p \times \mathbb{R}^{r+1}$. Then $f^A(\mathscr{X})$ corresponds to the expectation of the squared L_2 distance between the support functions of $\hat{\boldsymbol{\Theta}}$ and its expectation, see [16] for a study of this distance between convex sets. Hence, the A-optimal design aims to minimize the expected square distance between $\hat{\mathbf{\Theta}}$ and its expectation. The E-optimal design minimizes the L_∞ distance between the variance of the support function and zero, which is the support function of the origin. In case of the objective function (3) for the G-optimal design, the metric is mixed — the L_2 distance for one component of the tensor product $\mathbb{R}^p \times \mathbb{R}^{r+1}$ and the L_∞ distance for the other one.

In the classic linear univariate response setting with normal errors, the D-optimal design minimizes the confidence ellipsoid of parameter $\theta \in \mathbb{R}^{r+1}$

$$\{\theta : (\theta - \hat{\theta})^\top \Sigma^{-1} (\theta - \hat{\theta}) \le c\}$$

for a constant c, where $\hat{\theta}$ is the least square estimator of θ. The volume of this ellipsoid is proportional to $\det(\Sigma)^{-1/2}$, so that the objective function for D-optimal design becomes $\det(\Sigma)^{-1}$.

In the multiresponse setting of dimension p, it is possible to vectorize the parameter matrix by modifying the linear equation as

$$\begin{pmatrix} Y_1 \\ \\ Y_p \end{pmatrix} = \begin{pmatrix} \mathscr{X} & & \\ & \ddots & \\ & & \mathscr{X} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}. \tag{15}$$

The vector on the left-hand side arises by stacking together n observations for each component Y_j, $j = 1, \dots, p$, of the response. Furthermore, $\operatorname{diag}(\mathscr{X}, \dots, \mathscr{X})$ is an $np \times (r+1)p$ block-diagonal matrix built of the $n \times (r+1)$ dimensional design matrix $\mathscr{X}$; $\theta_j \in \mathbb{R}^{r+1}$, $j = 1, \dots, p$, are the estimated parameters, and ε_j, $j = 1, \dots, p$, are n-dimensional random vectors. Using the vector representation (15), Chang [4] proved that under the framework of approximate designs the D-optimal design in the multiresponse model is exactly the D-optimal design arising in the case of a univariate response. Kurotschka and Schwabe [11] extended this reduction result for both exact and approximate designs for D-, A-, E-optimality criteria and for more general Φ-optimality defined by Kiefer [8].

However, it is not possible to come up with an analogue of (15) for multivariate set-identified responses. In this case, one estimates the support function of $\mathbf{\Theta} = \mathbf{E}\hat{\mathbf{\Theta}}$, which is an infinite-dimensional parameter. Still, if one aims to reduce the integrated variance of $s(\hat{\mathbf{\Theta}}u, x)$ for each x, then the optimal design is the classical D-optimal one, see Remark 4.4.

It is worth to mention that our results in this paper are also applicable for multivariate polynomial regression models with set-identified response. In the classical setting of point-identified responses, Krafft and Schaefer [10] determined the approximate D-optimal design for the polynomial regression model and obtained a partial result for exact n-point D-optimal designs complemented later by Imhof [6] with a conjecture on G-optimum.

Our setting is restricted to the case of responses identified to belong to convex sets. In the non-convex setting, even the estimation of parameters is poorly understood not to mentioning the optimal design issues. In this case, the least square estimator (1) involves taking the sum of possibly non-convex sets. However, due to the convexification effect of Minkowski sums (see [12, Sec. 3.1.1]), the estimator $\hat{\mathbf{\Theta}}$ asymptotically becomes a convex set, and so such an estimator neglects the non-convexity of observations. Besides, we did not considered the case where the design matrix for each component of the response maybe different. This was thoroughly studied by Soumaya et al. [15] for the approximate D-optimal design with point-identified response.

Finally, note that our results are applicable only in the framework of [2]; the optimal design issues in the interval regression setting of [3] based on interval arithmetics do not fall into our scope of investigation.

Acknowledgment

The authors are grateful to the two anonymous referees for careful reading of the manuscript and supplying stimulating comments.

[1] A. C. Atkinson, A. N. Donev, and R. D. Tobias. *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford, 2007.

[2] A. Beresteanu and F. Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76:763–814, 2008.

[3] A. Blanco-Fernández, A. Colubi, and M. García-Bárzana. A set arithmetic-based linear regression model for modelling interval-valued responses through real-valued variables. *Inform. Sci.*, 247:109–122, 2013.

[4] S. I. Chang. Some properties of multiresponse D-optimal designs. *J. Math. Anal. Appl.*, 184:256–262, 1994.

[5] P. Diamond. Least squares fitting of compact set-valued data. *J. Math. Anal. Appl.*, 147:351–362, 1990.

[6] L. Imhof. Optimum designs for a multiresponse regression model. *J. Multivariate Anal.*, 72:120–131, 2000.

[7] M. Jacroux. On the MV-optimality of chemical balance weighing designs. *Calcutta Statist. Assoc. Bull.*, 32:143–151, 1983.

[8] J. Kiefer. General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, 2:849–879, 1974.

[9] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Can. J. Math.*, 12:363–366, 1960.

[10] O. Krafft and M. Schaefer. D-optimal designs for a multivariate regression model. *J. Multivariate Anal.*, 42:130–140, 1992.

[11] V. G. Kurotschka and R. Schwabe. The reduction of design problems for multivariate experiments to univariate possibilities and their limitations. In E. Brunner and M. Denker, editors, *Research Developments in Probability and Statistics*, pages 193–204. VSP, Utrecht, 1996.

[12] I. Molchanov. *Theory of Random Sets*. Springer, London, 2 edition, 2017.

[13] I. Molchanov and F. Molinari. *Random Sets in Econometrics*. Cambridge University Press, Cambridge, 2018.

[14] S. D. Silvey. *Optimal Design*. Chapman and Hall, London, 1980.

[15] M. Soumaya, N. Gaffke, and R. Schwabe. Optimal design for multivariate observations in seemingly unrelated linear models. *J. Multivariate Anal.*, 142:48–56, 2015.

[16] R. A. Vitale. L_p metrics for compact, convex sets. *J. Approx. Th.*, 45:280–287, 1985.